CLAIMS

What is claimed is:

1. A computer implemented method for prosody generation, comprising the steps of:

preparing an imitation speech database using recordings of natural human speech;

converting text to synthesized speech using a rule based speech synthesizer;

selecting prosody units from said imitation speech database to match said synthesized speech; and

concatenating said selected prosody units and generating a final prosody.

2. The method according to claim 1, wherein said rule based computer synthesizer uses a tone sequence prosody model.

3. The method according to claim 1, wherein said step of selecting prosody units from said imitation speech prosody database includes a cost function algorithm using distortion and concatenation costs.

4. The method according to claim 1, wherein said step of selecting prosody units from said imitation speech prosody database includes associating each syllable in said synthesized speech with an event including pitch events.

5.   The method according to claim 1, wherein said step of concatenating said selected prosody units and generating a final prosody includes an F0 smoothing function performed at concatenation points between selected prosody units.

6.   A computer implemented method for prosody generation, comprising the steps of:

preparing an imitation speech prosody database including:

converting training text to synthesized speech using a rule based computer synthesizer;

recording human speech imitating said synthesized speech;

time aligning said recorded human speech with said synthesized speech and extracting features from said recorded speech for syllables in which intonation events occur and generating said imitation speech prosody database; and

generating speech prosody from text including:

converting text to synthesized speech using a rule based synthesizer;

selecting prosody units from said imitation speech prosody database to match said synthesized speech; and

concatenating said selected prosody units and generating a final prosody.

7.   The method according to claim 6, wherein said rule based synthesizer uses a tone sequence prosody model.

8.   The method according to claim 6, wherein said step of selecting prosody units from said imitation speech prosody database includes a cost function algorithm using distortion and concatenation costs.

9.   The method according to claim 6, wherein said step of selecting prosody units from said imitation speech prosody database includes associating each syllable in said synthesized speech with an event including pitch events.

10.   The method according to claim 6, wherein said step of concatenating said selected prosody units and generating a final prosody includes an F0 smoothing function performed at concatenation points between selected prosody units.

11.   The method according to claim 6, wherein said step of time aligning said recorded human speech with said synthesized speech is performed using a dynamic time warp aligner.

12.   A speech generation processor for processing input text to speech, comprising:

an imitation speech database including prosodic units from imitation speech;

a rule based synthesizer module for generating synthesized speech curves for input text;

an imitation speech prosody selection module for selecting prosodic units from said imitation speech database with said synthesized speech curves and concatenating said selected prosodic units together for speech generation; and

an audible device for receiving a speech generation signal from said imitation speech prosody selection module and generating audible speech.